Quantification Precision

- Does your data-gathering instrument:
  - Provide consistent results?
  - Respond to change?
  - Provide a decision-aiding cut-point?
- Does your data-collection process reduce burden:
  - Of patient response?
  - Of provider administration?
  - Without losing information?

The New Measurement

- In the past 30+ years, measurement has undergone a quiet evolution
- Fundamental principles have been changed or even abandoned

A Rule That Has Broken

- To illustrate, focus on one particularly recognizable rule
- "Old" rule of interest: Longer tests are more reliable than shorter tests

The Context: Historical and Paradigmatic

- Beginnings of classical psychometrics go back to the turn of the 20th century

- Consider C. Spearman's work in disattenuating the correlation coefficient
  - *Demonstration of formulae for true measurement of correlation* (1907) American Journal of Psychology

The Zenith of Classical Psychometrics

- Arguably occurred in the time period surrounding the 1950 publication of H. Gulliksen's book, *Theory of Mental Tests*

- Book embodies many of the best and brightest contributions of classical psychometrics

Classical Test Theory (CTT)

- A cornerstone of classical psychometrics

- Is theory-based measurement
  - Individual scores are theory-defined as composed of a true score component and an error component
  - i.e., observed score = true score + error

An Issue with CTT

- CTT or "true score" theory does not provide
  - Hypotheses that are testable
  - Models that are falsifiable

Another Issue with CTT: Circular Definitions

- Item difficulty:
  - The proportion of examines in a group of interest who answer an item correctly
- Thus, an item being "hard" or "easy" depends on the ability of examinees measured
- Result: challenge in determining score meaning
  - Examinee and test characteristics are entangled
  - Each interpreted only in the context of the other

Modern Psychometrics

- In 1968 FM Lord and MR Novick's book, *Statistical Theories of Mental Tests*, was published
- Introduced model-based measurement, a foundation of modern psychometrics

Item Response Theory (IRT)

- IRT plays a fundamental role in modern psychometrics
- Addresses CTT's major shortcomings
  - By providing test-independent and group-independent measurement
  - By being model-based, employing models that can be tested and falsified
    - i.e., if a proposed IRT model does not adequately explain the data at hand, it is determinable that either assumptions were not met or an inappropriate model was used

CTT versus IRT

- CTT – classical psychometrics
  - Theory-based
  - Sample and test-specific
  - Focuses on test performance
- IRT – modern psychometrics
  - Model-based
  - Ability or functioning level-specific
  - Focuses on item performance

CTT Exercise

- Measurement instrument:
  - Lower Extremity Functional Scale (LEFS)
- Objective: to complete the LEFS "CTT" style, i.e.,
  - Provide answers for all activities
  - Sum responses
  - Interpret overall score

Original LEFS Instructions

We are interested in knowing whether you are having any difficulty at all with the activities listed below because of your lower limb problem for which you are currently seeking attention. Please provide an answer for each activity.

Today, do you or would you have any difficulty at all with:

(Circle one number on each line)

LEFS ITEMS “Activities”	Extreme difficulty or unable to perform activity	Quite a bit of difficulty	Moderate difficulty	A little bit of difficulty	No difficulty
1. Perform any of your usual work, housework, or school activities	0	1	2	3	4
2. Perform your usual hobbies, recreational or sporting activities	0	1	2	3	4
3. Getting into or out of the bath	0	1	2	3	4
4. Walking between rooms	0	1	2	3	4
5. Putting on your shoes or socks	0	1	2	3	4
6. Squatting	0	1	2	3	4
7. Lifting an object, like a bag of groceries from the floor	0	1	2	3	4
8. Performing light activities around your home	0	1	2	3	4
9. Performing heavy activities around your home	0	1	2	3	4
10. Getting into or out of a car	0	1	2	3	4
11. Walking 2 blocks	0	1	2	3	4
12. Walking a mile	0	1	2	3	4
13. Going up or down 10 stairs (about 1 flight of stairs)	0	1	2	3	4
14. Standing for 1 hour	0	1	2	3	4
15. Sitting for 1 hour	0	1	2	3	4
16. Running on even ground	0	1	2	3	4
17. Running on uneven ground	0	1	2	3	4
18. Making sharp turns while running fast	0	1	2	3	4
19. Hopping	0	1	2	3	4
20. Rolling over in bed	0	1	2	3	4

CTT Exercise Scenario

The patient:

- 62-year-old male
- Prior leg revascularization
- Current ABI measure: 0.80
- Experiencing some exertional calf pain that resolves within 10 minutes of rest
- Household activities, for the most part, not affected
- Unable to walk pet dog without some discomforts and multiple rest stops

CTT Exercise Products

- Products obtained:
  - Patient total score
  - Number of items didn't need to answer
  - Number of items needed to answer
- Also formulated ideas on:
  - Patient response burden
  - Provider administration burden

Response Burden

- Patient burden
  - Response time
  - Being asked inappropriate questions
- Provider burden
  - Score summation
  - Score interpretation
- What is the cost
  - Of patient burden?
  - Of provider burden?

Cost of Burden

- Cost of patient burden
  - Loss of willingness to respond in focused, honest way to instrument that seems unresponsive or even annoying
- Cost of provider burden
  - Loss of willingness to use instrument to collect data, score responses, or interpret patient self-reported condition

Burden Relief

- Computerizing test to relieve provider burden
  - Using scoring algorithm
  - Deliver score-specific interpretation
  - i.e., computer *administered* test
- What about patient burden?

IRT Exercise

- The LEFS as a computer *adaptive* test (CAT)
- Starting rule
  - Begin with an item of moderate difficulty
- Stopping rules
  - Three employed by LEFS CAT
    - When SEM < 4 (score range: 0-100)
    - When average of change for last 3 function estimates < 1
    - When complete all LEFS items

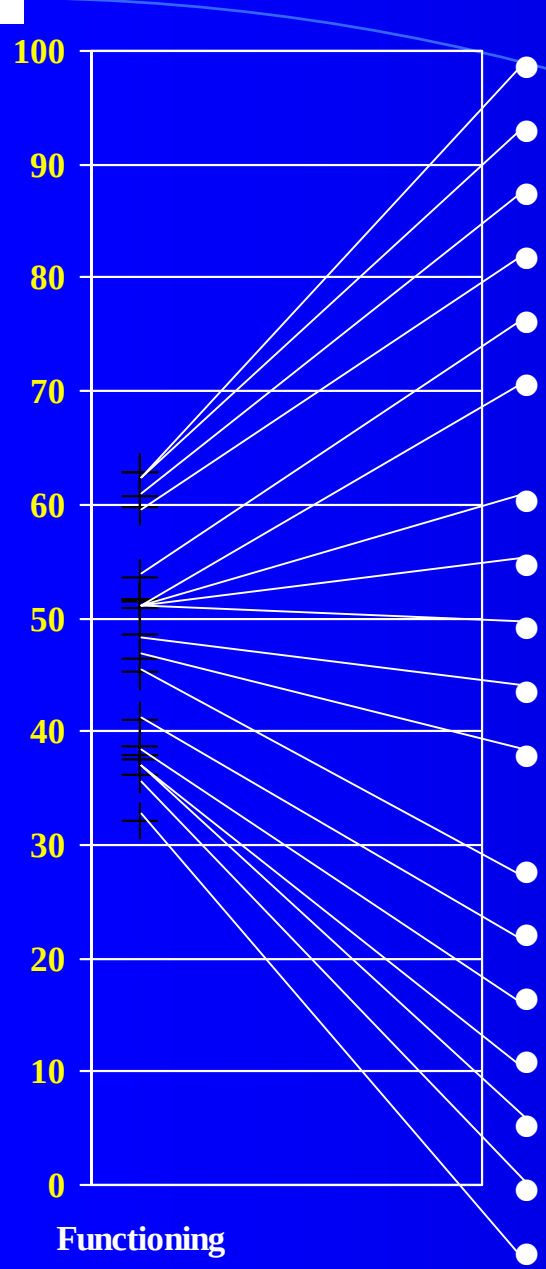
IRT Exercise Products

- Exercise products obtained:
  - Patient total score
  - Content of items needed to answer
  - Count of those items
  - Content of items didn't need to answer
  - Count of those items

LEFS: Item Structure

- Presentation order
  - CTT: standardized
    - All patients start at Item #1 and complete all LEFS items in order, going from #2 - #20
  - IRT: customized
    - Patients are presented new items based on their responses to previous items
- IRT "logic"
  - There is an underlying hierarchy of activities
  - Activities can be ordered ("calibrated") from hardest to easiest

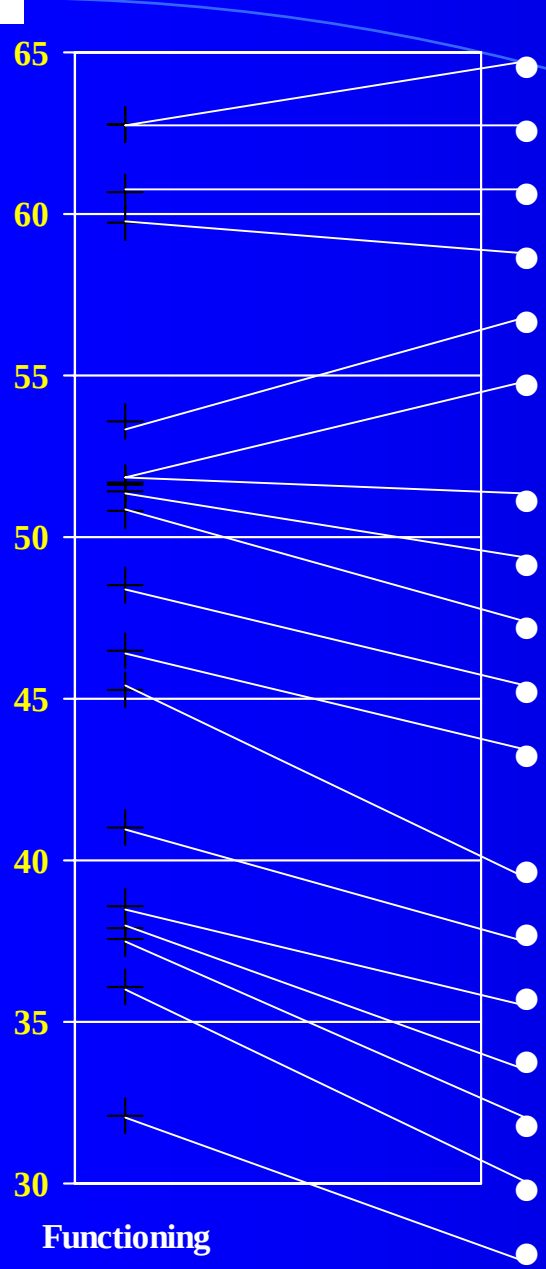
CALIBRATIONS	ITEMS
62.8 (hardest item)	Making sharp turns while running fast
62.8	Running on uneven ground
60.7	Running on even ground
59.7	Hopping
53.6	Walking a mile
51.7	Performing your usual hobbies, recreational or sporting activities
51.6	Squatting
51.4	Standing for 1 hour
50.8	Performing heavy activities around your home
48.5	Going up or down 10 stairs (about 1 flight of stairs)
46.5	Performing any of your usual work, housework, or school activities
45.3	Walking 2 blocks
41.0	Lifting an object, like a bag of groceries from the floor
38.6	Getting into or out of a car
37.9	Getting into or out of the bath
37.6	Performing light activities around your home
36.1	Putting on your shoes or socks
32.1 (easiest item)	Walking between rooms



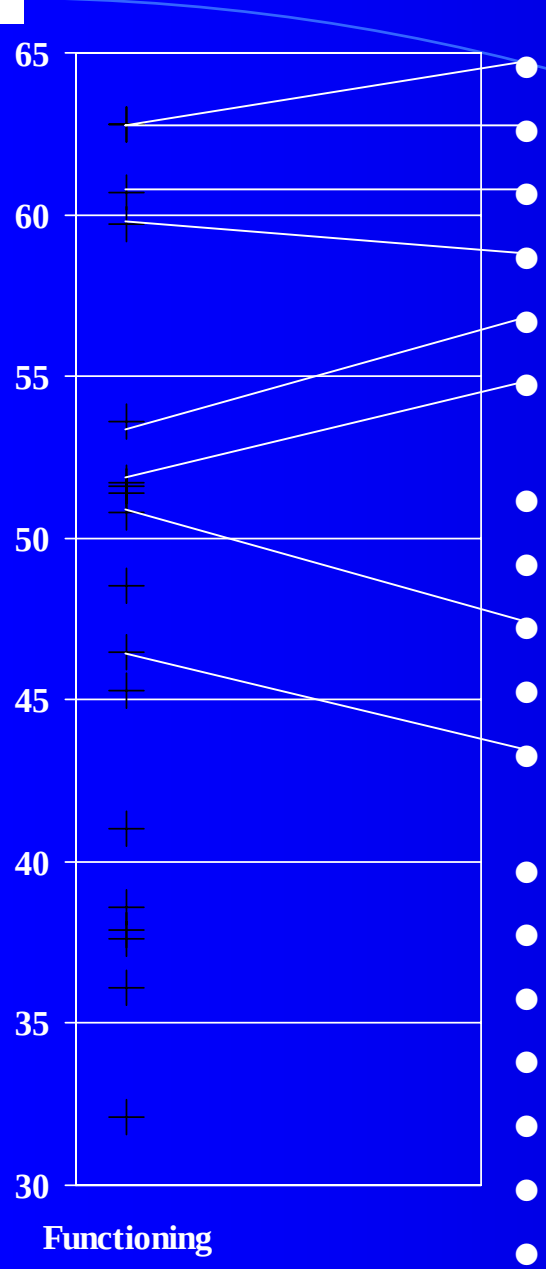
CALIBRATION: 0-100 SCALE

- Making sharp turns while running fast
- Running on uneven ground
- Running on even ground
- Hopping
- Walking a mile
- Performing your usual hobbies, recreational or sporting activities
- Squatting
- Standing for 1 hour
- Performing heavy activities around your home
- Going up or down 10 stairs (about 1 flight of stairs)
- Performing any of your usual work, housework, or school activities
- Walking 2 blocks
- Lifting an object, like a bag of groceries from the floor
- Getting into or out of a car
- Getting into or out of the bath
- Performing light activities around your home
- Putting on your shoes or socks
- Walking between rooms

CALIBRATION: 30-65 SCALE



- Making sharp turns while running fast
- Running on uneven ground
- Running on even ground
- Hopping
- Walking a mile
- Performing your usual hobbies, recreational or sporting activities
- Squatting
- Standing for 1 hour
- Performing heavy activities around your home
- Going up or down 10 stairs (about 1 flight of stairs)
- Performing any of your usual work, housework, or school activities
- Walking 2 blocks
- Lifting an object, like a bag of groceries from the floor
- Getting into or out of a car
- Getting into or out of the bath
- Performing light activities around your home
- Putting on your shoes or socks
- Walking between rooms



4. Making sharp turns while running fast
3. Running on uneven ground (0)
5. Running on even ground (0)
6. Hopping (1)
2. Walking a mile (1)
7. Performing your usual hobbies, recreational or sporting activities (2)
-
-
8. Performing heavy activities around your home (2)
-
1. Performing any of your usual work, housework, or school activities (3)
-
-
-
-
-
-
-

CALIBRATION: 30-65 SCALE

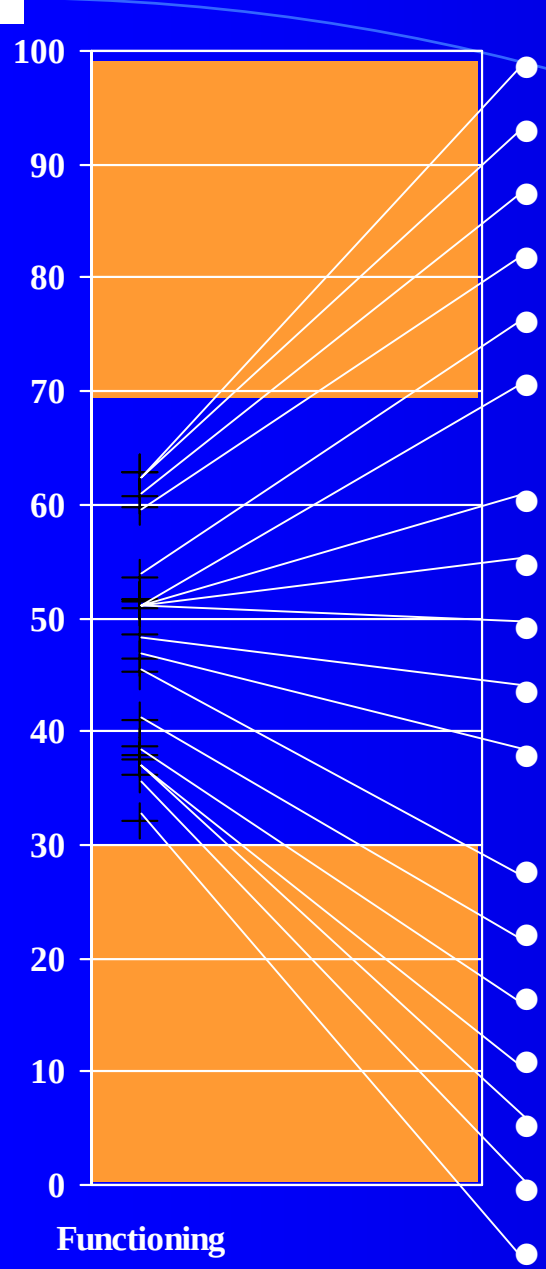
ITEM PRESENTATION ORDER AND RESPONSE FROM LEFS EXERCISE

Can the LEFS be improved?

- In what ways might the LEFS be improved?
- (Note: There are currently 18 items in LEFS CAT item bank)

Improving the LEFS

- Additional items can be created for areas on the functional scale that have no coverage
  - Extending the scale by adding items calibrated at >65 and <30



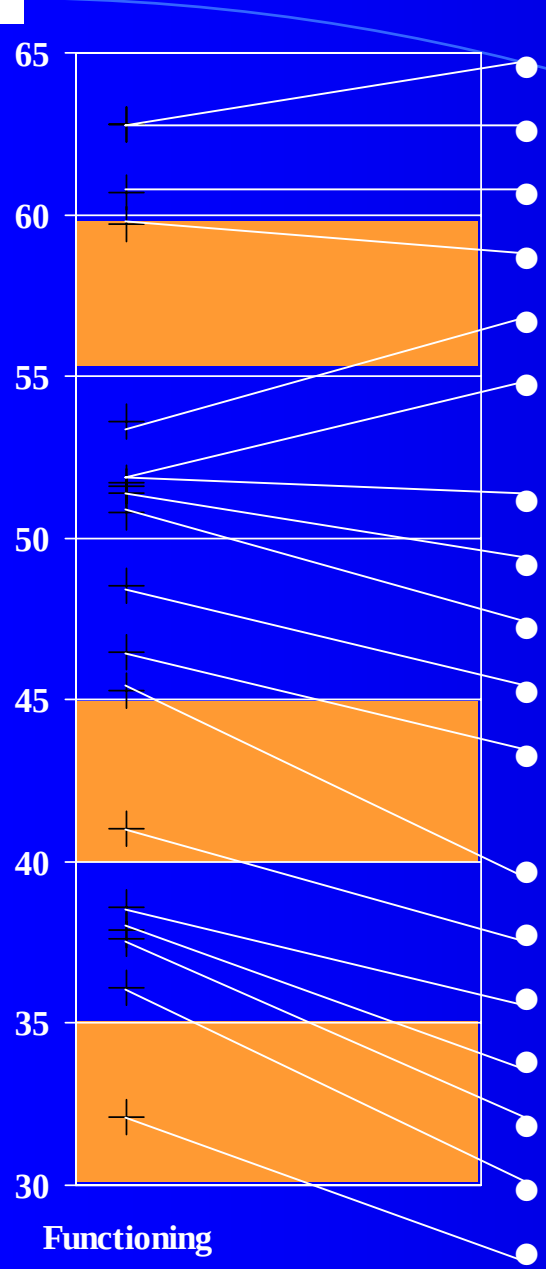
Making sharp turns while running fast
Running on uneven ground
Running on even ground
Hopping
Walking a mile
Performing your usual hobbies, recreational or sporting activities
Squatting
Standing for 1 hour
Performing heavy activities around your home
Going up or down 10 stairs (about 1 flight of stairs)
Performing any of your usual work, housework, or school activities
Walking 2 blocks
Lifting an object, like a bag of groceries from the floor
Getting into or out of a car
Getting into or out of the bath
Performing light activities around your home
Putting on your shoes or socks
Walking between rooms

CALIBRATION:
0-100 SCALE

COVERAGE AT >65 AND <30

Improving the LEFS

- Additional items can be created for areas on the functional scale that have poor coverage
  - Refining the scale by adding items in the coverage gaps between 30 and 65



Making sharp turns while running fast
Running on uneven ground
Running on even ground
Hopping
Walking a mile
Performing your usual hobbies, recreational or sporting activities
Squatting
Standing for 1 hour
Performing heavy activities around your home
Going up or down 10 stairs (about 1 flight of stairs)
Performing any of your usual work, housework, or school activities
Walking 2 blocks
Lifting an object, like a bag of groceries from the floor
Getting into or out of a car
Getting into or out of the bath
Performing light activities around your home
Putting on your shoes or socks
Walking between rooms

CALIBRATION: 30-65 SCALE

COVERAGE GAPS

Rules

“Old” Rule	“New” Rule
Longer tests are more reliable than shorter tests	Shorter tests can be more reliable than longer tests

Why aren't we all using CAT?

- Cost of computers and computing resources?
- Patient/provider unfamiliarity or discomfort with technology?
- Complexity of IRT modeling and calibration process?

Where to go from here?

- Read about IRT applications in the published literature
- Support IRT-based initiatives
- Use IRT methods in own clinical and research work
- Work together with other interested and capable parties

Future Prospects

The likelihood of soon seeing the wide availability of CATs in my area of practice?

- Modern measurement applications are of rapidly growing interest in medicine
- Technology currently exists to produce computer adaptive tests
- Therefore, it is expected that many new CATs will become commercially available in the near future